DOCUMENT RESUME

ED 435 693 TM 030 324

AUTHOR	Verhelst, N. D.; Kaftandjieva, F.
TITLE	A Rational Method To Determine Cutoff Scores. Research Report 99-07.
INSTITUTION	Twente Univ., Enschede (Netherlands). Faculty of Educational Science and Technology.
PUB DATE	1999-00-00
NOTE	18p.
AVAILABLE FROM	Faculty of Educational Science and Technology, University of Twente, TO/OMD, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Cutting Scores; Data Collection; Foreign Countries; *Item Response Theory; *Performance Based Assessment; *Standards
IDENTIFIERS	*Experts; *Standard Setting

ABSTRACT

A new method is proposed to set multiple standards in performance tests. The method combines three sources of information coming from three different data collections. The first is an empirical definition of mastery of an item; the second consists of parameter estimates of the items in an Item Response Theory (IRT) model, and the third source is a collection of experts' judgments on the relation between item mastery and level of performance. These judgments are given as an answer to very simple questions. The method is not iterative, and the experts are not required to judge borderline persons. The standard setting procedure is simple and can be carried out without a computer. (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 435 693

A Rational Method to Determine Cutoff Scores

Research Report 99-07

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

J. Nelissen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

N.D. Verhelst
Cito, Arnhem / University of Twente, Enschede

F. Kaftandjieva
University of Jiväskylä, Finland

TM030324

faculty of
EDUCATIONAL SCIENCE AND TECHNOLOGY

Department of Educational Measurement and Data Analysis

University of Twente

BEST COPY AVAILABLE

A Rational Method to Determine Cutoff Scores

N.D. Verhelst,
Cito, Arnhem /
University of Twente, Enschede

F. Kaftandjieva
University of Jiväskylä, Finland

Abstract

A new method is proposed to set multiple standards in performance tests. The method combines three sources of information coming from three different data collections. The first is an empirical definition of mastery of an item; the second consists of parameter estimates of the items in an IRT model, and the third source is a collection of experts judgements on the relation between item mastery and the level of performance. These judgments are given as an answer to very simple questions. The method is not iterative, and the experts are not required to judge borderline persons. The standard setting procedure is simple and can be carried out without computer.

Introduction

In this paper a procedure is developed to find multiple cutoff points on a scale. The framework of the procedure can be described as follows.

(1) The scale is described a priori in a number (R) of ordered levels, which are meant to cover the whole range of the proficiency being measured. Each level is described in rather general terms of performance.

(2) A number of items - larger than the number of categories- is constructed, administered to a sample, called calibration sample hereafter, and the responses are analyzed using a unidimensional IRT-model. The assumptions of the IRT-model are tested in an appropriate way, and possibly a number of items are discarded. It will be assumed in the sequel that the remaining items comply in a satisfactory way with the IRT-model used. Therefore the items together define a latent scale, and administration of the test to any person makes it possible to locate this person with known accuracy on the latent scale. The scale values will be symbolized by θ. The number of items will be denoted I

(3) J (> 1) experts in the subject field are given a training with the purpose to induce a quite homogeneous understanding of what is meant by the different levels of performance. Experts do not know the testees nor have ever seen any tables or statistics with information on the responses by the calibration sample.

(4) The experts give, after training, answers to $I(R-1)$ questions, phrased as: "Do you think a person at level r should should be able to answer this item (i) correctly?" ($r = 2, \ldots, R$; $i = 1, \ldots, I$). The experts have the text of the items at their disposal. The answers with regard to formerly presented levels are not available when answering a given level; so the experts cannot check their own consistency. The answers are binary (yes-no). The experts answer the questions independently of each other; so there is no discussion to reach agreement.

Notice important differences with existing classical standard setting procedures (see Berk, 1986, for an overview). The targeted population is not well defined, so that normative elements in determining the cutoff point do not enter the decision process. The main difference with important classical methods, however, is that there is no need to imagine a boderline person as in the methods of Angoff, Ebel and Nedelsky; reference is made (or intended to be made) to the concept of the level as induced in the training of the experts. As shall be come clear in the sequel, this vagueness can be a source of criticism to the procedure proposed, but it can also enhance the validity of the procedure. In this sense it is similar to Jaeger's procedure (Cross et al., 1984; Jaeger, 1993), but in contrast to this procedure, it is not iterative, and there is a slight

difference in the phrasing of the main question to the experts.

Although the method resembles the 'contrasting groups' method of Livingston and Zieky (1982), here there are important differences as well. The experts in the present method do not classify any person they know, where in the contrasting groups method a sample of persons having taken the test are classified. Moreover, in the latter method the judgment of the expert concerns a decision about the person, whereas in the present method the judgment aims at gathering information on the relation between typical persons and the items of the test.

The rationale for finding cutoff points rests on a comparison of the judgment of the experts and the characteristics of the item as found in the calibration sample. There is, however, a third component in the procedure which is essential, viz. the concept of mastery of an item. This will be explained in the next section.

Mastery of an item

The advantage of the instruction given to the experts is that it is free of any probabilistic element. At the same time, however, it is not said explicitly what is meant by the phrase 'should give a correct answer'. Does it mean a correct response always under any circumstance, or does it mean something like 'most of the time'? In other words, an explicit definition of mastery of an item is not given or induced. As will be seen in the sequel, lack of such a definition leaves the result of the procedure arbitrary to a certain extent. To find a unique solution, a definition of mastery has to be adopted.

A reasonable approach might be to define mastery in probabilistic terms: an item is mastered by a student if he has a probability of at least p to give a correct answer, where p is some number between zero and one. Of course, the outcome of the standard setting procedure will depend on the precise value of p, and therefore the choice of p should be founded on empirical evidence which reflects in some sense a widely accepted definition.

Such an empirical procedure might go as follows. A panel of experts expresses its concept of mastery for a test consisting of 50 parallel items, say, as a minimal proportion of correct responses. The average reported proportion - given there is not too much variability - can be taken as a measure of mastery. The important remark to make here, however, is that there is a procedure possible which is independent of the procedure described in the introductory section, and that it is not advisable to mix the two procedures. In the mastery definition procedure, the minimum requirement implicitly evokes the idea of a borderline person, while in the data collection described in the introduction no such reference is made

The loss function

To develop the argumention further, it will be assumed that the number p is fixed. As an example consider the situation where the items are calibrated using the two parameter logistic model (2PLM). The mastery level for item i is defined to be the minimal value of θ, such that the probability of a correct answer is at least p; this value will be denoted κ_i, and it is seen that it is the solution of the equation

$$\frac{\exp[\alpha_i(\kappa_i - \beta_i)]}{1 + \exp[\alpha_i(\kappa_i - \beta_i)]} = p, \tag{1}$$

where α_i and β_i are the known discrimination and difficulty parameters respectively. The solution is given by

$$\kappa_i = \beta_i + \frac{1}{\alpha_i} \ln \frac{p}{1-p}. \tag{2}$$

To find a rational cutoff point on the θ-axis which separates level $r+1$ from level r, $(r = 1, \ldots, R-1)$, only the expert judgements collected in the $r+1$-th condition will be used. The unknown boundary value between level r and level $r+1$ will be denoted x_r. Define the binary variable D_{ijr}, with realizations d_{ijr}, as taking the value one if rater j judges that a person of level r should give a correct response to item i, and zero otherwise.

The event $D_{ijr} = 1$ will be interpreted as the statement that all persons with a θ-value not smaller than x_r master item i. But this implies, by the reasoning above, that the judge makes a prediction about the relative positions of x_r and κ_i, stating that $\kappa_i \leq x_r$. Now, for some choice of x_r such that this relation does not hold, for all θ in the interval (x_r, κ_i), the prediction is wrong, and a positive loss is given which is a non-decreasing function of the length of the interval (x_r, κ_i).

For the case $D_{ijr} = 0$, the interpretation is not so clear. The negation of the statement given to the judge might mean: "it is not true that every student at level r should master this item", meaning that some do and some do not. It might also be interpreted that nobody at level r masters the item. This latter interpretation, however, leads to problems in case $r = R$, because it would imply that nobody ever can master that item. Therefore the former interpretation will be used, yielding, in terms of the latent scale, the prediction that $\kappa_i > x_r$. If the reverse relation holds, this prediction is wrong and a positive loss is associated with it, which is non-decreasing function of the length of the interval (κ_i, x_r).

Since finding the cutoff value is accomplished independently for each level, the subscript r referring to the level will be dropped from now on. Summarizing, the loss function in its most general form is defined as

$$L_{ij}(x) = d_{ij}g_1(|\kappa_i - x|)I(\kappa_i \geq x) + (1 - d_{ij})g_0(|\kappa_i - x|)I(\kappa_i \leq x), \tag{3}$$

where $g_1(.)$ and $g_0(.)$ are non-decreasing continuous functions over the non-negative reals, and $I(.)$ is the indicator function, taking the value 1 if its argument is true and 0 otherwise. Because a weak inequality is used in the indicator functions, it is reasonable to require that the equality $g_1(0) = g_0(0)$ holds, and without loss of generality we can require that $g_1(0) = g_0(0) = 0$, which ensures that the minimal loss is zero.

The overall loss function is defined as

$$L(x) = \sum_i w_i \sum_j v_j L_{ij}(x) \tag{4}$$

where w_i and v_j are fixed positive weights assigned to individual items and judges respectively.

Of course, there is a large number of possibilities in choosing the functions g_1 and g_0. A first choice occurs when a reason has to be found to choose different functions, reflecting that a wrong prediction should be penalized differently for different outcomes of the variable D_{ij}. In the present context there seems to be no reason for doing this, so we choose

$$g_1 = g_0 = g \tag{5}$$

Further considerations may concern continuity and differentiability of $L_{ij}(x)$. A very simple (and for many reasons attractive) function is given by

$$g(y) = \begin{cases} 1 \text{ if } y > 0 \\ 0 \text{ if } y = 0 \end{cases} \tag{6}$$

which makes L_{ij} a step function, jumping from 0 to 1 at κ_i if $D_{ij} = 1$ and the other way around if $D_{ij} = 0$. To construct a continuous function, we must ensure that $\lim_{y \to 0} g(y) = 0$. A class of functions which fulfill this requirement is given by

$$g(y) = y^k, \; k > 0 \tag{7}$$

If $k \leq 1$, the function L_{ij} is not differentiable at κ_i, but for $k > 1$, it is. So, choosing $k = 2$ yields a differentiable loss function with all the desirable characteristics that squared loss functions have in multivariate analysis. An important advantage will become clear in the sequel.

So, choosing $g_1 = g_0 = g$, and $g(y) = y^2$, (3) can be written as

$$L_{ij}(x) = \begin{cases} (\kappa_i - x)^2 \text{ if } D_{ij} = 1 \text{ and } \kappa_i > x, \\ (\kappa_i - x)^2 \text{ if } D_{ij} = 0 \text{ and } \kappa_i < x, \\ 0 \text{ otherwise.} \end{cases} \tag{8}$$

Of course, the definition of the overall loss function (4) remains the same. The optimal value of x is defined that value that minimizes the overall loss function. Although (8) looks simple enough, the minimization is not trivial: for given x the value of $L_{ij}(x)$ can assume only two different values. Which one applies, however, depends not only on the data D, but also on the value of x itself. Therefore, the loss function is not a simple quadratic function (with fixed coefficients), but it has variable coefficients. The procedure to minimize this function will be discussed in the next section.

Minimizing the loss function

Without loss of generality we can assume that the items are ordered in increasing order of their κ-value. Now, consider the closed interval $[\kappa_g, \kappa_{g+1}]$ for some $1 \leq g < I$. This interval will be referred to as the g-th interval in the sequel. For all values of x in this interval, the truth value of the two conditions given in the right hand side of (8) cannot change for any of the I items. Define the weights $f_i^{(g)}$, $(i = 1, \ldots, I;\ g = 1, \ldots, I-1)$ as

$$f_i^{(g)} = \begin{cases} w_i \sum_j v_j (1 - d_{ij}) \text{ if } i \leq g, \\ w_i \sum_j v_j d_{ij} \text{ if } i > g. \end{cases} \tag{9}$$

One of the following events must occur: either all weights are zero, or there is at least one positive weight. The latter case, which is the most interesting one, will be dealt with first.

The weight $f_i^{(g)}$ is the coefficient of the positive loss $(\kappa_i - x)^2$ in the overall loss function. If the restriction that x must be in the g-th interval is dropped, the overall loss function reduces to a simple quadratic function whose (unique) minimum - remind the advantage of choosing quadratic loss functions - is given by

$$y^{(g)} = \frac{\sum_i f_i^{(g)} \kappa_i}{\sum_i f_i^{(g)}}. \tag{10}$$

Restriction to the g-th interval gives immediately that the minimum of the loss function in this interval is attained at

$$m^{(g)} = \begin{cases} \kappa_g & \text{if} \quad y^{(g)} < \kappa_g, \\ \kappa_{g+1} & \text{if} \quad y^{(g)} > \kappa_{g+1}, \\ y^{(g)} & \text{if} \quad \kappa_g \leq y^{(g)} \leq \kappa_{g+1} \end{cases} \tag{11}$$

To complete the search, also the intervals $(-\infty, \kappa_1]$ and $[\kappa_I, +\infty)$ must be investigated. Consider the former interval, to be called the zero-th interval. Since $y^{(0)}$ is a weighted average of the κ's, it will necessarily follow that $y^{(0)} \geq \kappa_1$, and therefore

$$m^{(0)} = \kappa_1.$$

By a similar reasoning, it holds that

$$m^{(I)} = \kappa_I.$$

So the minimum of the overall loss function is at $m^{(h)}$ which is defined by

$$L(m^{(h)}) = \min_g L(m^{(g)}), \tag{12}$$

where g ranges from 1 to $I - 1$.

Of course this minimum exists, but it does not follow from the procedure described above that it is unique. There might be two local minima, which in some cases could be equal. To show that this is impossible, consider Figure 1 which shows the loss function in the neighborhood of a minimum. The function is piecewise quadratic, and is continuously differentiable everywhere. Therefore the loss function is a spline of degree two with knots at the κ-values. Notice, that since the graph of the function in any interval coincides with a parabola with a minimum, the function is convex everywhere. Now, if there are two local minima, there must be a (local) maximum in between, but this implies that the function should be concave in some region, which is not possible. Therefore the minimum is unique.

insert Figure 1about here

Next, the case is considered where all weights $f_i^{(g)}$ are zero. To see how this can happen, consider a hypothetical example with two judges, giving d_{ij} values as shown in Table 1, where the columns represent the items

Table 1. Hypothetical data

1	$\cdots$	i	$i+1$	$\cdots$	I
1	$\cdots$	1	0	$\cdots$	0
1	$\cdots$	1	0	$\cdots$	0

It is easy to see that for every x in the i-th interval there are no mastery points (κ-values) to the left of x such that any of the judges stated that non-mastery of the corresponding item is allowed, and similarly for mastery points to the right of x: mastery is not necessary according to both judges. So there will be zero penalty for each of the judgments, and the loss function attains its lower boundary for all values in the i-th interval. If this interval has positive length, it follows that the minimum is not unique.

Notice that since the items are ordered in increasing value of κ, and assuming that all κ-values are different from each other, the case of zero loss can only occur if each judge give a Guttman pattern of responses (i.e. a 1 never follows a 0), and if all these Guttman patterns are equal across judges. If there are ties in the κ-values, the definition of the Guttman pattern can be relaxed slightly: a 1 must be assigned to all items whose κ is less that the tied κ, and a 0 to all items with a larger κ. If this happens for all judges, the minimum of the loss function is at the tied κ, and of course is unique.

It will be clear from consideration of Table 1 that the interval with zero loss is unique. This interval may also be $(-\infty, \kappa_1]$ or $[\kappa_I, +\infty)$, corresponding to the cases where for all judges and all items $D_{ij} = 0$ or $D_{ij} = 1$ respectively.

Conclusion

A procedure has been proposed to find multiple cutoff points on a latent continuum which is defined by a number of items, the responses to which can be adequately modeled by a unidimensional IRT-model. To apply the procedure, three kinds of information must be available. We discuss these three in short.

The central concept in the whole procedure is mastery of an item. It is proposed to define this concept as the minimal latent ability required to have at least a probability p of a correct answer. An experiment was described on how the value of p might be determined. Once there is agreement on this value, the mastery criterion κ_i for each item is uniquely determined for all (parametric) models with increasing item response functions. For models with non-monotone item response functions, this issue may problematic. As far as achievement or attainment testing is considered, monotone item response function are the rule; in attitude measurement, the concept of mastery is probably not adequate. This is a serious limitation of the method proposed in this article.

Of course, to have an estimate of the κ-values, observations must be collected from a calibration sample from the target population. This may be quite a difficult problem if the target population is not known, as for example when the test is to be administered in the future via internet to whoever is interested in it. Updating the item parameter estimates and checking the validity of the IRT-model at regular intervals when new data come in seems the only possible way out from this problem.

Another aspect related to the calibration is the accuracy of the parameter estimates. In the development above, the κ-values are treated as constants, but of course there is some estimation error associated with them. Good practice may be to choose the weights w_i in (4) inversely proportional to the square of the standard error of the item parameter estimates. (or even better, to extend the procedure to incorporate the covariances as well).

The third kind of information is the expert judgment on mastery or non mastery of the items by a rather vaguely described person of a well described level of proficiency. The vague description is used to let the expert make full use of his own representation of the intended level of proficiency, without further qualification such as 'borderline' persons or even some concrete persons he happens to know. At the same time, asking probabilistic statements is avoided, which may be especially attractive for experts in domains with little mathematical thinking and experience, such as language testing, for example. But even teachers of mathematics find the yes/no method '(...)clearer and easier to use than the more traditional Angoff probability estimation procedure' (Impara and Plake, 1997).Of course, one might criticize the procedure in as much that it suggests that deterministic statements are asked for, like 'a person of level A should make this item always correct', meaning that a single error on many similar items would imply that this person cannot be of that level. But this would contradict common sense and educational practice, where it is not expected that the most excellent student should obtain the

maximum score on all the examinations to deserve the highest degree.

The proposed method integrates these three kinds of information in an easy way to arrive at a rational determination of the multiple cutoff points. Of course the validity of the whole procedure depends on the validity of the constituent parts. Not only the calibration results must be used to test the validity of the IRT-model, also the validity of the expert judgments is at stake, as well in the definition of the mastery level as in the phase of determining cutoff points. In the former case, there must not be too large variation in the probability statements; in the latter procedure, not too many deviations from Guttman patterns (see Table 1) must occur, and if only Guttman patterns occur (perfect intra-rater consistency), the location where the ones switch to zeros must not differ too much across judges (inter-rater consistency). If these requirements are violated to a large degree, the procedure to determine cutoff points may still be applied, possibly with different weights assigned to the judges, but the validity of the results may be questionable.

References

Berk, J.A. (1986). A Consumer's Guide to Setting Performance Standards on Criterion-Referenced Tests. *Review of Educational Research, 56,* 137-172.

Cross, L.H., Impara, J.C., Frary, R.B. & Jaeger, R.M. (1984). A Comparison of Three Methods for Establishing Minimum Standards on the National Teacher Examinations. *Journal of Educational Measurement, 21, 113-129.*

Impara, J.C. & Plake, B. (1997). Standard Setting: an Alternative Approach. *Journal of Educational Measurement, 34,* 353-366.

Jaeger, R.M. (1993). Certification of Student Competence. In: R.L. Linn (Ed.), *Educational Measurement (pp. 485-514).* Washington: American Council of Education.

Livingston, S.A. & Zieky, M.J. (1982). *Passing Sores: a Manual for Setting Standards of Performance on Educational and Occupational Tests.* Princeton, NJ: Educational Testing Service.

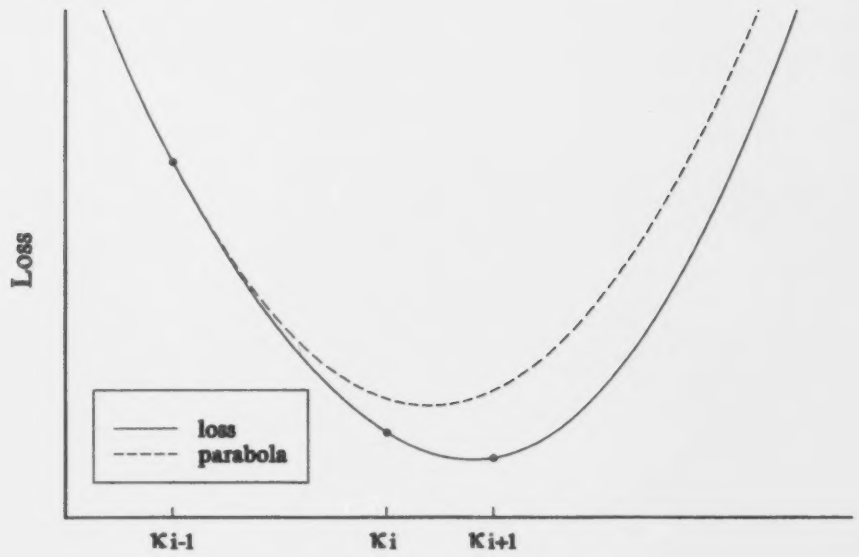

Figure 1.

The loss function and the parabola associated with the leftmost interval

Titles of Recent Research Reports from the Department of Educational Measurement and Data Analysis. University of Twente, Enschede, The Netherlands.

RR-99-07 N.D. Verhelst & F. Kaftandjieva, *A Rational Method to Determine Cutoff Scores*

RR-99-06 G. van Engelenburg, *Statistical Analysis for the Solomon Four-Group Design*

RR-99-05 E.M.L.A. van Krimpen-Stoop & R.R. Meijer, *CUSUM-Based Person-Fit Statistics for Adaptive Testing*

RR-99-04 H.J. Vos, *A Minimax Procedure in the Context of Sequential Mastery Testing*

RR-99-03 B.P. Veldkamp & W.J. van der Linden, *Designing Item Pools for Computerized Adaptive Testing*

RR-99-02 W.J. van der Linden, *Adaptive Testing with Equated Number-Correct Scoring*

RR-99-01 R.R. Meijer & K. Sijtsma, *A Review of Methods for Evaluating the Fit of Item Score Patterns on a Test*

RR-98-16 J.P. Fox & C.A.W. Glas, *Multi-level IRT with Measurement Error in the Predictor Variables*

RR-98-15 C.A.W. Glas & H.J. Vos, *Adaptive Mastery Testing Using the Rasch Model and Bayesian Sequential Decision Theory*

RR-98-14 A.A. Béguin & C.A.W. Glas, *MCMC Estimation of Multidimensional IRT Models*

RR-98-13 E.M.L.A. van Krimpen-Stoop & R.R. Meijer, *Person Fit based on Statistical Process Control in an AdaptiveTesting Environment*

RR-98-12 W.J. van der Linden, *Optimal Assembly of Tests with Item Sets*

RR-98-11 W.J. van der Linden, B.P. Veldkamp & L.M. Reese, *An Integer Programming Approach to Item Pool Design*

RR-98-10 W.J. van der Linden, *A Discussion of Some Methodological Issues in International Assessments*

RR-98-09 B.P. Veldkamp, *Multiple Objective Test Assembly Problems*

RR-98-08 B.P. Veldkamp, *Multidimensional Test Assembly Based on Lagrangian Relaxation Techniques*

RR-98-07 W.J. van der Linden & C.A.W. Glas, *Capitalization on Item Calibration Error in Adaptive Testing*

RR-98-06 W.J. van der Linden, D.J. Scrams & D.L.Schnipke, *Using Response-Time Constraints in Item Selection to Control for Differential Speededness in Computerized Adaptive Testing*

RR-98-05 W.J. van der Linden, *Optimal Assembly of Educational and Psychological Tests, with a Bibliography*

RR-98-04 C.A.W. Glas, *Modification Indices for the 2-PL and the Nominal Response Model*

RR-98-03 C.A.W. Glas, *Quality Control of On-line Calibration in Computerized Assessment*

RR-98-02 R.R. Meijer & E.M.L.A. van Krimpen-Stoop, *Simulating the Null Distribution of Person-Fit Statistics for Conventional and Adaptive Tests*

RR-98-01 C.A.W. Glas, R.R. Meijer, E.M.L.A. van Krimpen-Stoop, *Statistical Tests for Person Misfit in Computerized Adaptive Testing*

RR-97-07 H.J. Vos, *A Minimax Sequential Procedure in the Context of Computerized Adaptive Mastery Testing*

RR-97-06 H.J. Vos, *Applications of Bayesian Decision Theory to Sequential Mastery Testing*

RR-97-05 W.J. van der Linden & Richard M. Luecht, *Observed-Score Equating as a Test Assembly Problem*

RR-97-04 W.J. van der Linden & J.J. Adema, *Simultaneous Assembly of Multiple Test Forms*

RR-97-03 W.J. van der Linden, *Multidimensional Adaptive Testing with a Minimum Error-Variance Criterion*

RR-97-02 W.J. van der Linden, *A Procedure for Empirical Initialization of Adaptive Testing Algorithms*

RR-97-01 W.J. van der Linden & Lynda M. Reese, *A Model for Optimal Constrained Adaptive Testing*

RR-96-04 C.A.W. Glas & A.A. Béguin, *Appropriateness of IRT Observed Score Equating*

RR-96-03 C.A.W. Glas, *Testing the Generalized Partial Credit Model*

RR-96-02 C.A.W. Glas, *Detection of Differential Item Functioning using Lagrange Multiplier Tests*

RR-96-01 W.J. van der Linden, *Bayesian Item Selection Criteria for Adaptive Testing*

RR-95-03 W.J. van der Linden, *Assembling Tests for the Measurement of Multiple Abilities*

RR-95-02 W.J. van der Linden, *Stochastic Order in Dichotomous Item Response Models for Fixed Tests, Adaptive Tests, or Multiple Abilities*

RR-95-01 W.J. van der Linden, *Some decision theory for course placement*

RR-94-17 H.J. Vos, *A compensatory model for simultaneously setting cutting scores for selection-placement-mastery decisions*

...

Research Reports can be obtained at costs, Faculty of Educational Science and Technology, University of Twente, TO/OMD, P.O. Box 217, 7500 AE Enschede, The Netherlands.

faculty of

EDUCATIONAL SCIENCE AND TECHNOLOGY

A publication by
The Faculty of Educational Science and Technology of the University of Twente
P.O. Box 217
7500 AE Ens
The Netherla

BEST COPY AVAILABLE